2005

Stereophonic acoustic echo cancellation

Fenglu Ge

Prashan Premaratne
University of Wollongong, prashan@uow.edu.au

Jiangtao Xi
University of Wollongong, jiangtao@uow.edu.au

Stereophonic acoustic echo cancellation

Abstract

This paper presents a new approach for Stereophonic acoustic echo cancellation (SAEC), based on a combination of J-H network and the natural gradient algorithm. In the new approach for SAEC, Computer simulations show that echoes are significantly reduced.

Publication Details

Ge, F., Premaratne, P. & Xi, J. (2005). Stereophonic acoustic echo cancellation. In B. J. Wysocki & T. A. Wysocki (Eds.), 8th International Symposium on DSP and Communication Systems, DSPCS'2005 & 4th Workshop on the Internet, Telecommunications and Signal Processing, WITSP'2005 Noosa, Australia: DSP for communication Systems.

Stereophonic Acoustic Echo Cancellation

Fenglu Ge, Prashan Premaratne, Jiangtao Xi_
School of Electrical, Computer and Telecommunications Engineering
Faculty of Informatics, University of Wollongong
Wollongong, NSW 2522, Australia

flg5099@uow.edu.au

Abstract: **This paper presents a new approach for Stereophonic acoustic echo cancellation (SAEC), based on a combination of J-H network and the natural gradient algorithm. In the new approach for SAEC, Computer simulations show that echoes are significantly reduced.**

I. INTRODUCTION

Stereophonic acoustic echo cancellation (SAEC) [1] is an active research topic due to its applications in many fields including video teleconference, virtual reality, and so on. A major issue in Stereophonic acoustic echo cancellation is the non-uniqueness problem that results in misconvergence and slow convergence of adaptive filters in SAEC [1] [2]. As the non-uniqueness is mainly due to the high cross-correlation between two channels of stereo signal [1] [2], several pre-processing methods have been proposed to overcome the problem. The problem has been overcomed by trying to reduce the correlation between the two channels, such as the approaches based on nonlinear transformation [2] [3] [4] [5], the methods of input moving technique [6] [7], and the techniques using beamforming [8] [9] [10]. However, all of these approaches have some weakness. For example, nonlinear transformation and input moving technique suffer from degradation of sound quality, while for beamforming, the echo cancellation system need a complex training procedure first, and the training position should be desired [10].

In order to achieve SAEC without losing the sound quality, a post-processing method is introduced by Okuno &Tokhi(2001) [11]. Unlike the pre-processing technique mentioned above, post-processing technique is better in that the far-end signals are sent to the near-end room without any distortion. However, the approach proposed by Okuno &Tokhi might not work well as the far-end signal is assumed to contain two uncorrelated speech signals. This is not what happens in practice. Also in [11] the convergence is too slow due to the use of the INFOMAX algorithm [12].

In this paper we propose a new post-processing technique for SAEC without including the double-talk problem. In the next section, the proposed solution is described and analyzed. In Section III, the simulation results of the proposed scheme are given. In Section IV, the paper is concluded with summative comments regarding the success and future problem of SAEC..

II. THE PROPOSED APPROACH

The post-processing approach is depicted in Fig.1.The stereo signal consisting of two channels, left channel signal $x_1(n)$ and right channel signal $x_2(n)$, are sent from the far-end in the left box to the near-end. n is the time index for discrete signals. Due to acoustic coupling between speakers and microphones, echoes are picked up by the two microphones:

$$d_1(n) = h_{11}(n) * x_1(n) + h_{21}(n) * x_2(n) \tag{1}$$

$$d_2(n) = h_{21}(n) * x_1(n) + h_{22}(n) * x_2(n) \tag{2}$$

where $d_1(n)$ and $d_2(n)$ are signals collected by the two microphones respectively, and $h_{11}(n)$ $h_{12}(n)$ $h_{21}(n)$ and $h_{22}(n)$ are the room acoustic impulse responses from speakers to the microphones. * denotes convolution.

As $x_1(n)$ and $x_2(n)$ are from the same stereo signal, they are highly correlated, and hence from Equations (1) and (2) the two echoes $d_1(n)$ and $d_2(n)$ must also be highly correlated. Conventional echo cancellers using adaptive filters are not able to achieve good performance due to the non-uniqueness problem [1]. In order to solve the problem we use blind source separation (BSS) to reduce the correlation between the two echoes and then use adaptive echo cancellers to eliminate the echoes.

In this paper, we use J-H network [13] for BSS, which is shown in Fig.2.

$$d_{22}(n)=d_2(n)+\sum_{k=1}^{N}d_{11}(n-k+1)w_{1k2}(n) \tag{4}$$

where N denotes the number of filter coefficients and $w_{ikj}(n)$ is the k-th filter coefficient of the filter W_{ij} at

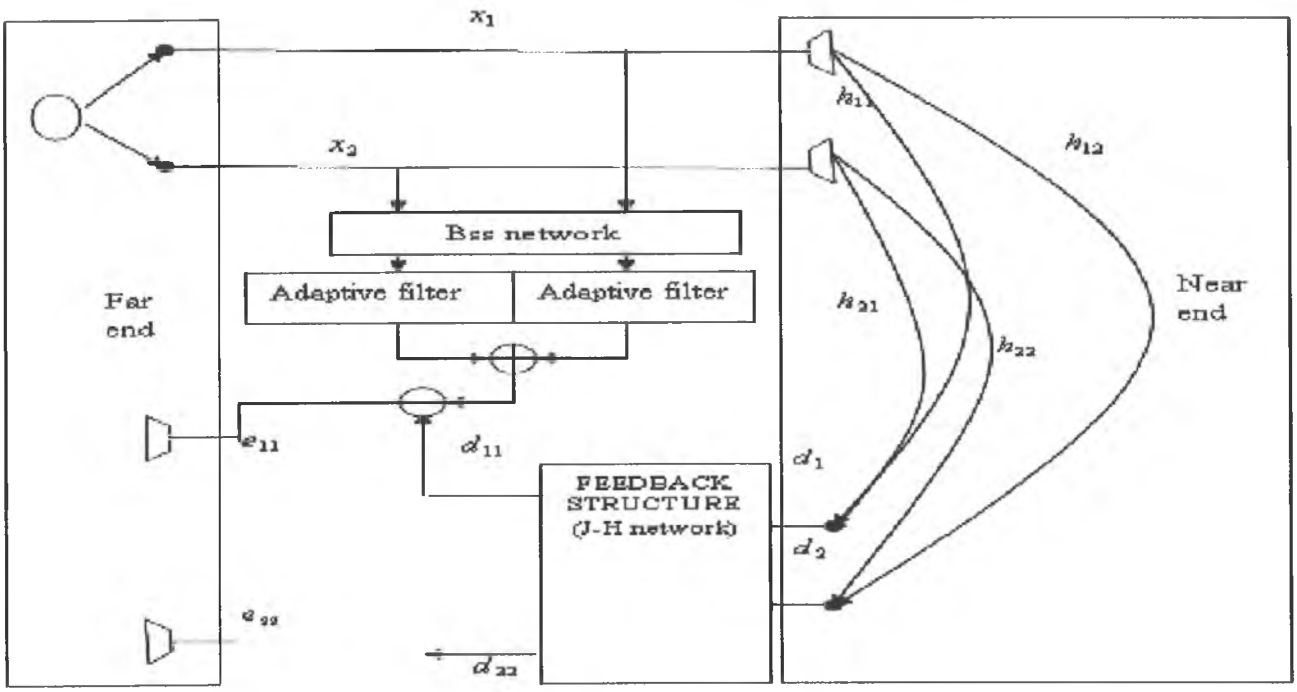

Fig.1. Scheme of proposed system on post-processing method

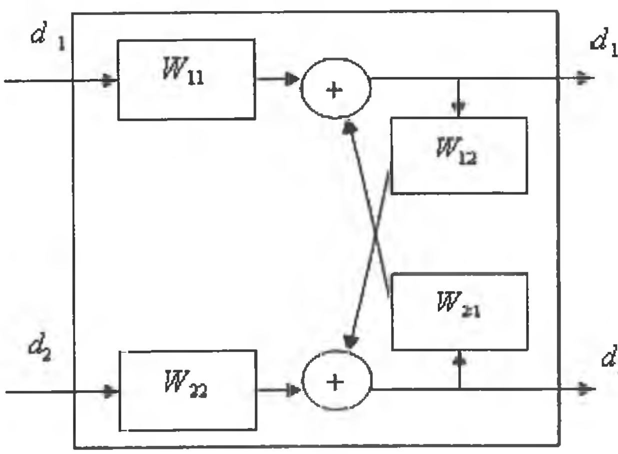

Fig.2. Scheme of J-H network

As shown in Fig.2, the relations between the input and output of J-H network are as follows:

$$d_{11}(n)=d_1(n)+\sum_{k=1}^{N}d_{22}(n-k+1)w_{2k1}(n) \tag{3}$$

time n ,and $w_{11}(n)$ and $w_{22}(n)$ are set to be kronecker delta separately.

As all signals are from the same source, it is obvious that $d_{11}(n)$ and $d_{22}(n)$ are still correlated. However, the correlation between $d_{11}(n)$ and $d_{22}(n)$ must be weaker compared to that between $d_1(n)$ and $d_2(n)$, which will make the echo cancellation more effective.

The signals $x_1(n)$ and $x_2(n)$ are also correlated because they are from the same sources.

To cancel $d_{11}(n)$ and $d_{22}(n)$, the signals generated from the BSS network are used in the adaptive filters. The BSS network is used to decorrelate $x_1(n)$ and $x_2(n)$, which is shown as follows:

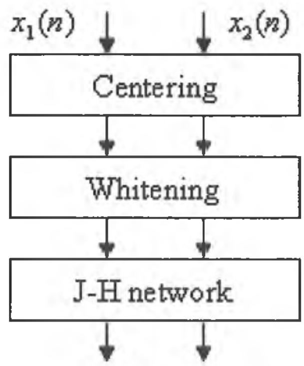

Fig.3.Scheme of BSS network.

Fig.3 can be summarized as three procedures:

1) The input vector $[x_1(n), x_2(n)]$ is centered, which means subtracting their mean values [13].

2) The signals from the centering model are whitened to reduce computation complexity [13].

3) The signals from the whitening model go through the J-H network to become little correlated signals.

The decorrelated $x_1(n)$ and $x_2(n)$ are defined as $\hat{x}_1(n)$ and $\hat{x}_2(n)$.

Since $\hat{x}_1(n)$ and $\hat{x}_2(n)$, $d_{11}(n)$ and $d_{22}(n)$ are from the same source and decorrelated by the same algorithm. It can be understood that either $\hat{x}_1(n)$ or $\hat{x}_2(n)$ is highly correlated with $d_{11}(n)$, while remaining one is little correlated with $d_{11}(n)$. Therefore, most of $d_{11}(n)$ can be eliminated by either $\hat{x}_1(n)$ or $\hat{x}_2(n)$ which are highly correlated with $d_{11}(n)$. With adaptive filter, residual part of $d_{11}(n)$ can be eliminated by another signal with adaptive filter. And the same structure can be applied to $d_{22}(n)$.

Finally, the NLMS algorithm is applied to adaptive filters to eliminate acoustic echoes.

Another issue is the optimization technique used for the J-H network. In order to achieve better convergence, we employ the natural gradient algorithm [14] [15] [16].

The natural gradient algorithm employed can be summarized as following:.

True transfer function has a relation with the matrix w, where

$$w(n) = \begin{bmatrix} w_{11} & w_{12}(n) \\ w_{21}(n) & w_{22} \end{bmatrix} \quad (5)$$

$$w_{12}(n) = [w_{112}(n), w_{122}(n), \cdots w_{1N2}(n)]^T \quad (6)$$

$$w_{21}(n) = [w_{211}(n), w_{212}(n), \cdots w_{2N1}(n)]' \quad (7)$$

It is obvious that w is not a square matrix but a $2N \times 2$ matrix. In the Riemannian space, for the nontrivial value of w, it can be achieved [16]

$$< dw, dw >_w \approx < dwy, dwy >_{wy} \quad (8)$$

where $\approx$ means approximation and $<,>$ means inner product. If $wy = I$, which means the space composed by w and y is Euclidean space, it shows:

$$< dw, dw >_w \approx < dwy, dwy >_{wy} = y^T (dw)^T (dw) y \quad (9)$$

and the Riemannian metric is $pinv(w)^T \, pinv(w)$, where $pinv$ means Moore-Penrose pseudo inverse and T means transposing. Then the gradient function between Riemannian space and Euclidean space is:

$$\nabla R \approx \nabla E \times [pinv(w)^T \times pinv(w)]^{-1} \quad (10)$$

where ∇R means the gradient in the Riemannian space, and ∇E means the gradient in the Euclidean space.

With the natural gradient algorithm, the adaptation rule for coefficients of filters W_{ij} is derived as follows:

$$\nabla w_{11}(n) = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{N\times 1} = \mu_1 \frac{\partial \ln|J|}{\partial w_{21}} [w_{112}(n) + w_{211}(n)] \quad (11)$$

$$\nabla w_{12}(n) = \begin{bmatrix} \nabla w_{112}(n) \\ \vdots \\ \nabla w_{1N2}(n) \end{bmatrix}_{N\times 1}$$

$$=\mu_2\begin{bmatrix}\frac{\partial \ln|J|}{\partial w_{112}(n)}\\ \vdots \\ \frac{\partial \ln|J|}{\partial w_{1N2}(n)}\end{bmatrix}\left[w_{112}(n)^2+w_{122}(n)^2\cdots+w_{1N2}(n)^2+1\right] \quad (12)$$

$$\nabla w_{21}(n)=\begin{bmatrix}\nabla w_{211}(n)\\ \vdots \\ \nabla w_{2N1}(n)\end{bmatrix}_{N\times 1}$$

$$=\mu_3\begin{bmatrix}\frac{\partial \ln|J|}{\partial w_{211}(n)}\\ \vdots \\ \frac{\partial \ln|J|}{\partial w_{2N1}(n)}\end{bmatrix}\left[w_{211}(n)^2+w_{221}(n)^2\cdots+w_{2N1}(n)^2+1\right] \quad (13)$$

$$\nabla w_{22}(n)=\begin{bmatrix}0\\ \vdots \\ 0\end{bmatrix}_{N\times 1}=\mu_4\frac{\partial \ln|J|}{\partial w_{12}}[w_{112}(n)+w_{211}(n)] \quad (14)$$

where ∇ denotes gradient, and J denotes a cost function.

$$|J|=\frac{\partial y_1}{\partial d_{11}}\frac{\partial y_2}{\partial d_{22}}-\frac{\partial y_2}{\partial d_{11}}\frac{\partial y_1}{\partial d_{22}} \quad (15)$$

$$y_1=\frac{1}{1+e^{-d_{11}}} \quad (16)$$

$$y_2=\frac{1}{1+e^{-d_{22}}} \quad (17)$$

μ_1, μ_2, μ_3 and μ_4 are coefficients added to increase system's stability and speed the convergence.

III. SIMULATION

Computer simulations were performed using Matlab on the basis of proposed approach, using the parameters in Table 1 and room impulse responses $h_{11}(n)$ $h_{12}(n)$ $h_{21}(n)$ and $h_{22}(n)$ in Fig.4.

A speech signal was used as the excitation signal. We use Echo Return Loss Enhancement ($ERLE$) to evaluate the performance of the proposal approach, which is defined as

$$ERLE=10\log_{10}\frac{E[y(k)^2]}{E[e(k)^2]} \quad (18)$$

E is defined as mathematical expectation. $y(k)$ is the signal received by the microphone in the near-end, which can be regarded as $d_1(n)$ or $d_2(n)$. $e(k)$ is the error signal, which denotes $e_{11}(n)$ or $e_{22}(n)$.

Fig.5 shows the Echo Return Loss Enhancement ($ERLE$) of the system, and no noise is added to the excitation signal. It can be concluded that, after post-processing, the echoes are significantly reduced.

Fig.6 shows the total $ERLE$ difference between method in [11] and proposed solution. It indicates the Echo Return Loss is enhanced by proposed solution.

total $ERLE=ERLE_{11}+ERLE_{22}$

where, $ERLE_{11}$ is the $ERLE$ in left channel and $ERLE_{22}$ is the $ERLE$ in right channel.

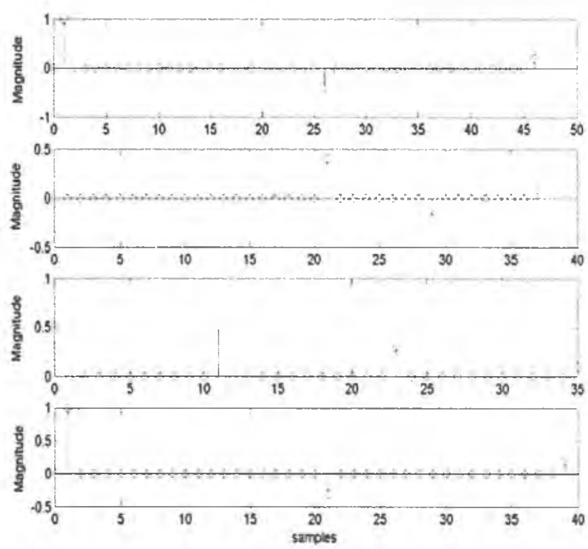

Fig.4. Room impulse responses in near-end

system's interference to near-end speech signals should be investigated to be reduced to a reasonable criterion.

TABLE I

PARAMETERS USED IN COMPUTER SIMULATION

Sampling frequency	8192Hz
Step-size in NLMS	0.01
$\mu_1, \mu_2, \mu_3\ \ \mu_4$	1.0e-9
Filter length in NLMS	28 Samples
Filter length in BSS network	2 Samples
Filter length in feedback structure	30 Samples

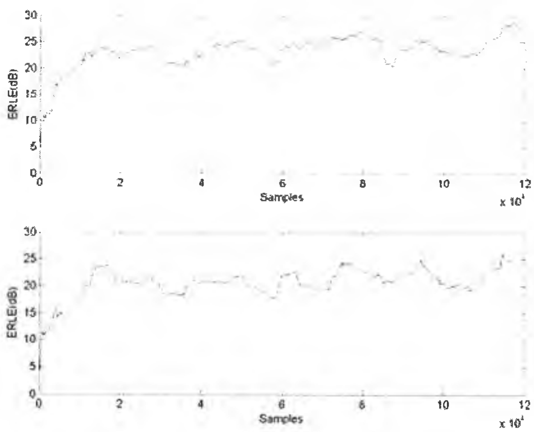

Fig.5. $ERLE$ of two channels h_{11} and h_{22}

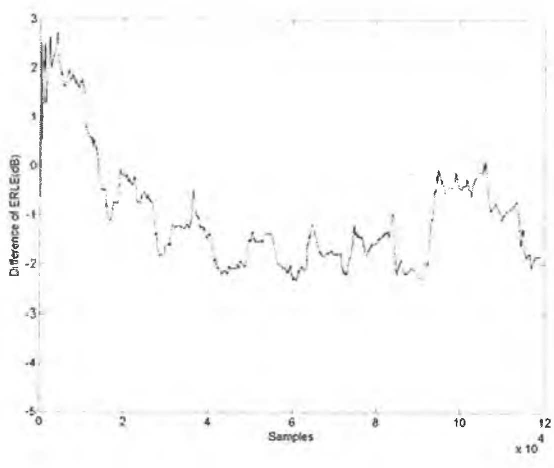

Fig.6. Difference of $ERLE$

IV. CONCLUSION

SAEC combining with the BSS algorithm is introduced in this paper. The simulation results show the system works well by post-processing the echoes. However, there is much work to do in the future. The convergence speed should be improved for real-time application, and the system's interference to near-end speech signals should be investigated to be reduced to a reasonable criterion.

ACKNOWLEDGMENT

The author would like to thank JiangTao Xi for patient reviewing and supplying many helpful suggestions.

REFERENCES

[1] M.M.Sondhi; D.R. Morgan; J.L. Hall, "Stereophonic acoustic echo cancellation-an overview of the fundamental problem," *IEEE. Signal Processing Letters*, vol.2, no.8, pp.148 – 151, Aug. 1995.

[2] J. Benesty; F. Amand; A. Gilloire; Y. Grenier, "Adaptive filtering algorithms for stereophonic acoustic echo cancellation," *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference*, vol.5, 9-12 pp.3099 – 3102, May 1995.

[3] D. R. Morgan; J. Benesty ; Sondhi, M.M.,"A better understanding and an improved solution to the specific problems of SAEC," *IEEE Tran. Speech and Audio Processing, vol.6*, no.2, pp.156 – 165,March 1998.

[4] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Trans. Speech and Audio Processing,* vol.9, no.6, pp.686-696, Sep 2001.

[5] Gansler, T.; Benesty, J,"New insights into the stereophonic acoustic echo cancellation problem and an adaptive nonlinearity solution," *IEEE Trans. Speech and Audio Processing,* Volume: 10, Issue: 5, July 2002 Pages: 257 - 267

[6] A. Sugiyama; Y. Joncour; A. Hirano; "A stereo echo canceller with correct echo-path identification based on an input-sliding technique", *IEEE Trans. Signal Processing,*, vol.49, no.11, pp. 2577 – 2587,Nov. 2001.

[7] A.Sugiyama; A.Hirano; K.Nakayama; "Acoustic Echo Cancellation For Conference Systems"

[8] See, C.M.S.; "Method for array calibration in high-resolution sensor array processing", *IEE Proceedings. Radar, Sonar and Navigation,*, vol.142, no.3, pp.90 - 96 June 1995.

[9] Leshem, A. Wax, M. RAFAEL, Haifa, "Array calibration in the presence of multipath", *IEEE Trans. Signal Processing,* pp. 53-59, Jan 2000.

[10] S. Nordholm; I.Claesson; M. Dahl, "Adaptive microphone array employing calibration signals: an analytical evaluation," *IEEE Trans. Speech and Audio Processing,* vol.7, no. 3, pp. 241 – 252, May 1999.

[11] T.Okuno,M.O.Tokhi, "Stereophonic acoustic echo cancellation using blind source separation as a post processing", *Proceedings of IWAENC-2001*,pp.75-78,2001

[12] A.J.Bell and T.J.Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution,"*Neural Computation,* vol.7,no.6,pp.1129-1159,1995

[13]A.Hyvärinen,E.Oja,"Independent component analysis: algorithms and applications", *Neural Networks*, v.13 n.4-5, p.411-430, May-June 2000.

[14] Herault, C.J.J, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetictecture." *Signal Processing*, vol.24, pp. 1-10, 1991.

[15] Amari, S.D., S.C., "Why natural gradient?" *Acoustics, Speech, and Signal Processing*, vol.2, pp. 1213-1216,1998.

[16] S.Amari, A.C., and H.H.Yang,"a new learning algorithm for blind signal separation." *Advances in Neural Information Processing Systems*, 1996. 8.

[17] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," 11th IFAC Symp. Syst. Ident., Kitakyushu, Japan, July 1997.

[18] K. Torkkola," Blind separation of convolved sources based information maximization", IEEE Workshop on Neural Networks for Signal Processing, pp.423-432,1996.

[19] S.Amari, "Natural gradient works efficiently in learning", *Neural Computation*, vol.10, pp.251-276, 1998